Citation for published version:
Morters, P & Jacob, E 2013, A spatial preferential attachment model with local clustering. in *Algorithms and Models for the Web Graph.: Proceedings of the 10th International Workshop, WAW 2013, Cambridge, MA, USA.*. vol. 8305, Lecture Notes in Computer Science, Springer, pp. 14-25. https://doi.org/10.1007/978-3-319-03536-9_2

DOI:
10.1007/978-3-319-03536-9_2

Publication date:
2013

Document Version
Early version, also known as pre-print

Link to publication

University of Bath

Alternative formats
If you require this document in an alternative format, please contact:
openaccess@bath.ac.uk

General rights
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy
If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Download date: 05. May. 2023

A spatial preferential attachment model with local clustering

Emmanuel Jacob and Peter Mörters

[1] École Normale Supérieure de Lyon
[2] University of Bath

Abstract. A class of growing networks is introduced in which new nodes are given a spatial position and are connected to existing nodes with a probability mechanism favouring short distances and high degrees. The competition of preferential attachment and spatial clustering gives this model a range of interesting properties. Most notably, empirical degree distributions converge to a limit law, which can be a power law with any exponent $\tau > 2$, and the average clustering coefficient converges to a positive limit. Our main tool to show these and other results is a weak law of large numbers in the spirit of Penrose and Yukich, which can be applied thanks to a novel rescaling idea. We also conjecture that the networks have a robust giant component if τ is sufficiently small.

Keywords: Scale-free network, Barabasi-Albert model, preferential attachment, dynamical random graph, geometric random graph, power law, degree distribution, edge length distribution, clustering coefficient.

1 Introduction

Many of the phenomena in the complex world in which we live have a rough description as a large network of interacting components. It is therefore a fundamental problem to derive the global structure of such networks from basic local principles. A well established principle is the *preferential attachment paradigm* which suggests that networks are built by adding nodes and links successively, in such a way that new nodes prefer to be connected to existing nodes if they have a high degree [3]. The preferential attachment paradigm offers, for example, a credible explanation of the observation that many real networks have degree distributions following a *power law behaviour*. On the global scale preferential attachment networks are *robust* under random attack if the power law exponent is sufficiently small, and have logarithmic or doubly logarithmic diameters depending on the power law exponent. These features, together with a reasonable degree of mathematical tractability, have all contributed to the enormous popularity of these models. Unfortunately, the local structure of preferential attachment networks significantly deviates from that observed in real networks. In preferential attachment models the neighbourhoods of typical nodes have a tree-like topology [11], [4], which is a crucial feature for their mathematical analysis, but is not in line with the behaviour of many real world networks.

The most popular quantity to measure the local clustering of networks are the *clustering coefficients*, which are measured to be positive in most real networks, but which invariably vanish in preferential attachment models that do not incorporate further effects [2], [6]. A possible reason for the clustering of real networks is the presence of a hidden variable assigned to the nodes, such that similarity of values is a further incentive to form links. For the class of protean graphs this idea has allowed Bonato et al. in [7] to generate power law networks with spatial clustering. Several authors have also proposed models combining preferential attachment with spatial features. Among the mathematically sound attempts in this direction are the papers of Flaxman et al. [12], [13], Jordan [16], Aiello et al. [1] and Cooper et al. [8]. These papers show that combining preferential attachment and spatial dependence can retain the global power law behaviour while changing the local topology of the network, for example by showing that the resulting graphs have small separators [12], [13]. None of these papers discusses clustering by analysing the clustering coefficients.

In this paper, we study a generalisation and variant of the spatial preferred attachment (SPA) model introduced in Aiello et al. [1] and further studied in Janssen et al. [15] and Cooper et al. [8]. The original model is based on the idea that a vertex at position x has a ball of influence centred in x. A new vertex can only be connected to it, if it falls within this ball, in which case it does so with a given probability p_0. The preferential attachment effect is modelled through the fact that the size of the ball depends on the degree of the vertex. In our model, this ball of influence is replaced by a profile, rotationally symmetric around x, with the probability of a connection given by the height of the profile. This allows us to relax the spatial rigidity of the model, so that for example two vertices always have a positive probability of being connected, whatever their positions. This generalisation induces a richer phenomenology, in particular when it comes to more complex statistics such as the edge length distribution or the existence of a giant component.

Our analysis of this model is using methods developed originally for the study of random geometric graphs, see Penrose and Yukich [19] for a seminal paper in this area and [18] for an exhibition. Our approach is based on a rescaling which transforms the growth in time into a growth in space. This transformation stabilises the neighbourhoods of a typical vertex and allows us to observe convergence of the local neighbourhoods of typical vertices in the graph to an infinite graph. This infinite graph, which is not a tree, is locally finite and can be described by means of a Poisson point process. We establish a weak law of large numbers, similar to the one given in [19], which allows us to deduce convergence results for a large class of functionals of the graph. For example, we show that the average clustering coefficient always converges to a positive constant for the scale-free networks given by SPA models. We also observe interesting phase transitions in the behaviour of the global clustering coefficient and the empirical edge length distribution. Finally, we informally discuss the existence of a robust giant component, one of the key features of preferential attachment networks which we would like to see retained in our model.

2 The model

The generalized SPA model may be defined in a variety of metric spaces S. In this paper, we work in dimension $d \geq 1$, and we choose a distance d on $\mathbb{R}^d$ derived from any of the l_p norms. Similarly as in [8], [15] we let S be the unit hypercube in $\mathbb{R}^d$, centred at 0, equipped with its own torus metric d_1, i.e. for any two points $(x, y) \in S$, we set $\mathsf{d}_1(x, y) = \min\{\mathsf{d}(x, y + u) : \ u \in \{-1, 0, 1\}^d\}$. Note that S equipped with the torus metric has no boundary and is spatially homogeneous, which avoids some technical difficulties. Let $\mathcal{X}$ denote a Poisson point process of unit intensity on $S \times (0, \infty)$. A point $\mathbf{x} = (x, s)$ in $\mathcal{X}$ is a vertex $\mathbf{x}$, born at time s and placed at position x. Observe that, almost surely, two points of $\mathcal{X}$ neither have the same birth time nor the same position. We say that (x, s) is *older* than (y, t) if $s < t$. An edge is always oriented from the younger to the older vertex. For $t > 0$, write $\mathcal{X}_t$ for $\mathcal{X} \cap (S \times (0, t])$, the set of vertices already born at time t. We construct a growing sequence of graphs $(G_t)_{t>0}$, starting from the empty graph, and adding successively the vertices in $\mathcal{X}$ when they are born (so that the vertex set of G_t is $\mathcal{X}_t$), and connecting them to some of the older vertices. The rule is as follows:

> **Construction rule**: Given the graph G_{t-} and $\mathbf{y} = (y, t) \in \mathcal{X}$, we add the vertex $\mathbf{y}$ and, independently for each vertex $\mathbf{x} = (x, s)$ in G_{t-}, we insert the edge $(\mathbf{y}, \mathbf{x})$, independently of $\mathcal{X}$, with probability
>
> $$\varphi\left(\frac{t^{1/d}\mathsf{d}_1(x, y)}{f(\deg^-(\mathbf{x}, t-))^{1/d}}\right). \tag{1}$$
>
> The resulting graph is denoted by G_t.

Here the following definitions and conventions apply:

(i) $\deg^-(\mathbf{x}, t-)$ (resp. $\deg^-(\mathbf{x}, t)$) denotes the indegree of vertex $\mathbf{x}$ at time $t-$ (respectively t), that is, the total number of incoming edges for the vertex $\mathbf{x}$ in G_{t-} (resp. G_t). Similarly, we denote by $\deg^+(\mathbf{y})$ the outdegree of vertex $\mathbf{y}$, which remains the same at all times $u \geq t$.

(ii) $f\colon \mathbb{N} \cup \{0\} \to (0, \infty)$ is the *attachment rule.* Informally, $f(k)$ quantifies the preferential 'strength' of a vertex of current indegree k, or likelihood of attracting new links. For simplicity, we suppose

$$f(k) = \gamma k + \beta, \quad \gamma \in (0, 1), \beta > 0,$$

just as in [1], [8], but most of the results hold unchanged if f is only supposed to be increasing with asymptotic slope $\lim_{k\to\infty} f(k)/k = \gamma$.

(iii) $\varphi\colon [0, \infty) \to [0, 1]$ is the *profile function.* It is non-increasing and satisfies

$$\int_{\mathbb{R}^d} \varphi\big(\mathsf{d}(0, y)\big)\, \mathrm{d}y = 1. \tag{2}$$

Informally, the profile function describes the spatial dependence of the probability that the newborn vertex $\mathbf{y}$ is linked to the existing vertex $\mathbf{x}$.

Loosely speaking, this form of the construction rule is the only one that ensures that we have a genuine interaction of the spatial and the preferential attachment effects, as a vertex is likely to be connected to a finite number of vertices within distance of order $t^{-1/d}$ and indegree of order 1, as well as to a finite number of vertices at distance $\gg t^{-1/d}$ and indegree $\gg 1$. If φ is integrable, the condition (2) is no loss of generality, as otherwise one can modify φ and f without changing the construction rule. Under (2) one can see that the interesting range of f (leading to degree distributions following an approximate power law) is characterised by asymptotic linearity. If $\gamma = pA_1$, $\beta = pA_2$ and $\varphi = p\mathbb{1}_{[0,r]}$, where r is chosen so that (2) is satisfied, we essentially get the original SPA model of [1], with the slight modification that we work in continuous time with random birth times. With this choice of profile function, the model can be interpreted as follows: Each vertex $\mathbf{x}$ is surrounded by a ball of influence, a ball centered at x and of volume $f(\deg^-(\mathbf{x}, t-))/(pt)$. If the new vertex $\mathbf{y}$ falls within this ball of influence, then $\mathbf{y}$ and $\mathbf{x}$ are connected with probability p, otherwise they cannot be connected.

A general profile function can be seen as a mixture of indicator functions, where any values $p \in (0, 1]$ are allowed. We are particularly interested in the case of profile functions φ with support the whole $\mathbb{R}_+$, in which case two vertices $\mathbf{x}$ and $\mathbf{y}$ always have a positive probability of being connected. In particular, we will discuss the choice of a polynomially decaying profile function, that is

$$\varphi(x) \asymp (1 + x)^{-\delta}, \quad \delta > d,$$

where $g \asymp h$ is the commonly used notation for g/h bounded away from zero and infinity. The condition $\delta > d$ is needed for the integrability condition to be satisfied.

We now illustrate the connection between non-spatial preferential and spatial attachment models. Suppose the graph G_{t-} is given, and a vertex is born at time t, but we do not know its position, which is therefore uniform on S_1. Then, for each vertex $\mathbf{x} = (x, s) \in G_{t-}$, the probability that it is linked to the newborn vertex is equal to

$$\int_{S_1} \varphi\left(K\mathsf{d}_1(x, y)\right) \mathrm{d}y = K^{-d} \int_{(-\frac{K}{2}, \frac{K}{2}]^d} \varphi(\mathsf{d}(0, y))\, \mathrm{d}y,$$

where we have written $K = t^{1/d}/(f(\deg^-(\mathbf{x}, t-))^{1/d})$. As a consequence, the *indegree evolution* process $(\deg^-(\mathbf{x}, t))_{t \geq s}$ is a time-inhomogeneous pure birth process, starting from 0 and jumping at time t from state k to state $k + 1$ with intensity

$$\frac{f(k)}{t} \int_{\left(-\frac{t^{1/d}}{2f(k)^{1/d}}, \frac{t^{1/d}}{2f(k)^{1/d}}\right]^d} \varphi(\mathsf{d}(0, y))\, \mathrm{d}y.$$

We can show that $(\deg^-(\mathbf{x}, t))_{t \geq s}$ grows roughly like t^γ, so that the integral is asymptotically 1. Hence the jumping intensity of our process is the same as in the Barabási-Albert model of preferential attachment [3], [20], or its variant studied by Dereich and Mörters [9], [10], [11].

As soon as one deepens the study of the graph further than the first moment calculations, the essential difference with the non-spatial models appears. The presence of edges is now strongly correlated through the spatial positions of the vertices. These correlations both make the model much harder to study, and allow the network to enjoy interesting clustering properties. These are the main concern of this paper.

3 Rescaling the graph

This section has been simplified from the full-version of this article, see [14]. The interested reader will find in [14] all the details, and the proof of Proposition 1 and Theorem 1 in the one-dimensional case. Everything holds *mutatis mutandis* in the higher dimensional cases

3.1 The rescaled picture

In the graph sequence $(G_t)_{t>0}$, the degree of any given vertex goes almost surely to $+\infty$. In this section we introduce a different graph sequence $(G^t)_{t>0}$ such that for every fixed t the graphs G_t and G^t have the same law. The new sequence has a different dynamics in which growth in time is replaced by growth in space, and the degrees of fixed vertices remain finite. Loosely speaking the sequence $(G^t)_{t>0}$ represents the graphs as seen from a typical vertex in the original graph sequence $(G_t)_{t>0}$, and hence a fixed point in $(G^t)_{t>0}$ does not age whereas a fixed point in $(G_t)_{t>0}$ does. The graph sequence $(G^t)_{t>0}$ will be easier to analyse, in particular it will converge and this goes along with convergence results for a large class of statistics derived from $(G_t)_{t>0}$.

To be more precise, let $\mathcal{Y}$ be a Poisson point process with intensity 1 on $\mathbb{R}^d \times (0,1]$. We interpret the first coordinate as space and the second as time, which is now restricted to the unit interval. For $t > 0$, we define S_t to be the hypercube

$$S_t = \left(-\frac{t^{1/d}}{2}, \frac{t^{1/d}}{2}\right]^d$$

of volume t. It is seen as a subspace of $\mathbb{R}^d$ but it is endowed with its own torus distance d_t. Observe that for any x and y of $\mathbb{R}^d$, for t large enough, x and y will be in S_t and satisfy $\mathsf{d}_t(x,y) = \mathsf{d}(x,y)$. For $t > 0$, define $\mathcal{Y}_t = \mathcal{Y} \cap (S_t \times (0,1])$, and construct a graph G^t on $\mathcal{Y}_t$ with the same construction rule as before, with the new understanding that time now belongs to $(0,1]$, and the distance is now replaced by d_t in (1). It is easily seen that the graphs G^t and the original graph G_t have the same law. Just multiply the time coordinate by t^{-1}, the space coordinates by $t^{1/d}$, and observe that the point process is still a Poisson point process of intensity 1, while the construction rule (1) is unchanged. It will turn out that there is a limiting graph $G^\infty = \lim G^t$, which can be obtained directly by applying our construction to the point set $\mathcal{Y}$ endowed with the distances in $\mathbb{R}^d$ in the construction rule (1).

3.2 Convergence

Proposition 1.

(i) The graph G^∞ is almost surely a well-defined locally finite graph, in the sense that each vertex has finite degree.
(ii) Almost surely, the graph G^t converges locally to G^∞, in the sense that for each $\mathbf{x} \in \mathcal{Y}$, for sufficiently large t, the neighbours of $\mathbf{x}$ in G^t and in G^∞ coincide.

The proposition states local convergence of G^t to G^∞ around any given vertex $\mathbf{x} \in \mathbb{R}^d \times (0,1]$. Its proof is based on a study of the indegree evolution process and bounds on the probability that a vertex has an exceptionally high indegree, or outdegree, or connects to an exceptionally distant vertex.

The following theorem completes Proposition 1 by describing the local structure of G^t around a *randomly chosen* vertex $\mathbf{x} \in G^t$. It is also the key to proving global results for the graphs G_t, see the following sections. It can be seen as a geometric law of large numbers in the spirit of Yukich and Penrose, the proof using that distant regions of space are asymptotically independent. For $t \in (0, \infty]$ let U be uniform on $(0,1]$ and $G^t(U)$ be the graph obtained by adding the point $(0, U)$ to $\mathcal{Y}$ before the construction of the graph G^t. Let $\xi(\mathbf{x}, G)$ be a 'local' function on a graph G around a distinguished point $\mathbf{x}$. For the purpose of this article, we can simply define such a local function to be a function on the neighbourhood of $\mathbf{x}$ up to graph distance a given finite value.

Theorem 1 (Weak law of large numbers). *Suppose, for some $a > 1$, the following uniform moment condition holds,*

$$\sup_{t>0} \mathbb{E}[\xi((0,U), G^t(U))^a] < \infty.$$

Then the following convergence in probability is satisfied,

$$\frac{1}{|\mathcal{Y}_t|} \sum_{\mathbf{x} \in \mathcal{Y}_t} \xi(\mathbf{x}, G^t) \longrightarrow \mathbb{E}[\xi((0,U), G^\infty(U))]. \tag{3}$$

In other words, the law of the local structure of the graph G^t around a randomly chosen vertex is the same as the law of the local structure of the infinite graph G^∞, conditioned[3] to have a vertex with position 0 and birth time U uniform in $(0,1]$, around this vertex. The next sections provide various applications to Proposition 1 and Theorem 1.

4 Results

Indegree. Denote by μ the law of the indegree of $(0,U)$ in $G^\infty(U)$ defined by

$$\mu(k) = \mathbb{P}(\deg^-((0,U), G^\infty(U)) = k).$$

[3] We recall here that for a Poisson point process, adding a point at $(0,U)$ is equivalent to conditioning the Poisson point process on having a point at $(0,U)$.

The local finiteness of $G^\infty(U)$ ensures that it is a probability law on $\mathbb{N} \cup \{0\}$. Applying the law of large numbers to the functionals $\xi_k(\mathbf{x}, G) = \mathbb{1}\{\deg^-(\mathbf{x}, G) = k\}$, we get that the empirical indegree distribution of G_t, defined by

$$\mu_t(k) = \frac{1}{|\mathcal{X}_t|} \sum_{\mathbf{x} \in \mathcal{X}_t} \mathbb{1}\{\deg^-(\mathbf{x}, G_t) = k\},$$

converges to μ in probability. Given the construction of $G^\infty(U)$, it is remarkable that the law of the indegree μ can be calculated explicitly. It relies on the study of the *indegree evolution* process, which we omit here. We find

$$\mu(k) = \frac{1}{\gamma} \frac{\Gamma(k + \frac{\beta}{\gamma})\Gamma(\frac{\beta+1}{\gamma})}{\Gamma(k + \frac{\beta+\gamma+1}{\gamma})\Gamma(\frac{\beta}{\gamma})} \sim \frac{\Gamma(\frac{\beta+1}{\gamma})}{\gamma\Gamma(\frac{\beta}{\gamma})} k^{-(1+1/\gamma)} \qquad \text{as } k \uparrow \infty,$$

which is in line with Theorem 1.1 of [1], and verifies the scale-free property of the network with power law exponent $\tau = 1 + 1/\gamma \in (2, \infty)$.

Actually, we can prove a stronger convergence result:

Theorem 2. *For any nondecreasing function* $g\colon \mathbb{N} \cup \{0\} \to [0, \infty)$*, we have the following convergence in probability, as* $t \to \infty$*,*

$$\sum \mu_t(k) g(k) \longrightarrow \sum \mu(k) g(k).$$

Applying this result with $g(k) = k$, we get that the total number of edges is always asymptotically of the same order as the number of vertices. More interestingly, applying it with $g(k) = k^2$, we get that

$$\frac{1}{|\mathcal{X}_t|} \sum_{\mathbf{x} \in \mathcal{X}_t} \deg^-(\mathbf{x}, G_t)^2$$

converges to a finite constant if $\gamma < 1/2$ and to infinity if $\gamma \geq 1/2$.

Outdegree and total degree. Similarly to the empirical indegree distribution we define the empirical outdegree distribution ν_t of G_t, and let ν the law of the outdegree of $(0, U)$ in $G^\infty(U)$. As before, Theorem 1 yields convergence in probability of ν_t to ν. We do not have an explicit expression for ν, in particular it is *not* a Poisson distribution as in the model of Dereich-Mörters [9]. By a study of the infinite picture, we can however prove that ν is light-tailed and get the following theorem.

Theorem 3. *For any* $0 < \alpha < 1 - \gamma$*, we have*

$$\nu([k, +\infty)) = o(e^{-k^\alpha}).$$

Moreover, for any function $g\colon \mathbb{N} \cup \{0\} \to \mathbb{R}$ *satisfying* $g(k) = o(e^{k^\alpha})$ *for some* $0 < \alpha < 1 - \gamma$*, we have the following convergence in probability, as* $t \to \infty$*,*

$$\sum \nu_t(k) g(k) \longrightarrow \sum \nu(k) g(k).$$

These results complete Theorem 1.5 in [1], which controls the maximum degree in G_t. Further, it is not hard to see that the law of the outdegree of $(0,u)$ in $G^\infty(u)$ is independent of u, and that the law of the total degree of $(0,U)$ in $G^\infty(U)$ is the convolution $\mu * \nu$. Hence, the empirical total degree distribution in G_t converges to $\mu * \nu$, which is also decaying polynomially with parameter τ. We can check that the mean degree is

$$\sum k\ \mu * \nu(k) = 2\sum k\nu(k) = 2\sum k\mu(k) = \frac{2\beta}{1-\gamma}.$$

Clustering. In this section, we forget the orientation of the edges of G_t to define its clustering coefficients. These coefficients are based on the number of triangles and open triangles in the graph. An open triangle of G_t with tip x is simply a subgraph of the form $(\{\mathbf{x},\mathbf{y},\mathbf{z}\},\{\{\mathbf{x},\mathbf{y}\},\{\mathbf{x},\mathbf{z}\}\})$, where $\mathbf{y}$ and $\mathbf{z}$ could either be connected in G_t and hence form a triangle, or not. Note that every triangle in G contributes three open triangles.

The *global clustering coefficient* of G is defined as

$$c^{\text{glob}}(G_t) := 3\,\frac{\text{Number of triangles included in } G_t}{\text{Number of open triangles included in } G_t}.$$

Note that always $c^{\text{glob}}(G_t) \in [0,1]$. The local clustering coefficient of G_t at a vertex $\mathbf{x}$ with degree at least two is defined by

$$c^{\text{loc}}_{\mathbf{x}}(G_t) := \frac{\text{Number of triangles included in } G_t \text{ containing vertex } \mathbf{x}}{\text{Number of open triangles with tip } \mathbf{x} \text{ included in } G_t},$$

which is also an element of $[0,1]$. Finally, the *average clustering coefficient* is defined as

$$c^{\text{av}}(G) := \frac{1}{|V_2|}\sum_{\mathbf{x}\in V_2} c^{\text{loc}}_{\mathbf{x}}(G),$$

where $V_2 \subset V$ is the set of vertices with degree at least two in G. The global and average clustering coefficients have the following probabilistic interpretation:

- *Pick a vertex uniformly at random and condition on the event that this vertex has degree at least two. Pick two of its neighbours, uniformly at random. Then the probability that these two vertices are linked is equal to $c^{\text{av}}(G_t)$.*
- *Pick two edges sharing a vertex, uniformly from all such pairs of edges in the graph. Then the probability that the two other vertices bounding the edges are connected is equal to $c^{\text{glob}}(G_t)$.*

Theorem 4.

(i) The average clustering coefficient of G_t converges in probability to a strictly positive number c^{av}_∞.

(ii) The global clustering coefficient of G_t converges in probability to nonnegative number c^{glob}_∞, which is nonzero if and only if $\gamma < 1/2$.

The first part of this theorem is easy to prove by considering the functional which, to a vertex, associates its local clustering coefficient. It is clear that the expected local clustering coefficient of $(0, U)$ in $G^\infty(U)$ belongs to $(0,1)$, and there is nothing more to argue. For the second part, we estimate both the number of triangles and the number of open triangles in G_t thanks to two applications of Theorem 1. We choose to count the triangles *from their youngest vertex* and consider the functional which associates to a vertex $\mathbf{x}$ the number of triangles containing $\mathbf{x}$ and having $\mathbf{x}$ as youngest vertex. The light-tail of the outdegree distribution ensures that this functional satisfies the uniform moment condition. We can apply Theorem 1 and deduce that the number of triangles is always asymptotically proportional to the number of vertices, that is of order t. To estimate the number of open triangles, we should in particular estimate the number of open triangles with tip in $\mathbf{x}$ the *oldest vertex*. But this number is simply

$$\sum_{\mathbf{x}\in\mathcal{X}_t} \deg^-(\mathbf{x}, G_t)(\deg^-(\mathbf{x}, G_t) - 1),$$

and we know, thanks to the work on the indegree, that it is linear if $\gamma < 1/2$ and superlinear if $\gamma \geq 1/2$. This discussion is almost enough to prove Theorem 4.

An interesting extension, suggested by an anonymous referee, is to look at the average local clustering coefficient of vertices with a fixed degree k. Our methods are expected to show that this quantity converges to a deterministic limit, which decays of order $1/k$, as $k \to \infty$. Details will be discussed elsewhere.

The phase transition in the global clustering coefficient has been observed in a similar form for random intersection graphs [5]. The behaviour of the clustering coefficients in the case $\gamma \geq 1/2$ matches the behaviour expected in the world wide web: if you pick a webpage at random, and click on two hyperlinks, it is likely that the two pages you get have actually a direct hyperlink. However, if you pick two webpages which both have a hyperlink to the Google homepage, it is not likely that these two pages have a direct link.

Edge length distribution. In the rescaled graphs G^∞ or in G^t, we expect a typical edge to have geometric length (in $\mathbb{R}^d$ or in S_t) of order 1. Therefore, in the original graph G_t, we expect edges to have length of order $t^{-1/d}$. Write $E(G_t)$ for the set of the edges of the graph G_t. Define λ, the (rescaled) empirical edge length distribution, by

$$\lambda_t = \frac{1}{|E(G_t)|} \sum_{(\mathbf{x},\mathbf{y})\in E(G_t)} \delta_{t^{1/d}\mathtt{d}_1(\mathbf{x},\mathbf{y})}.$$

Similarly, write $E(G^\infty(U))$ for the set of the edges of the graph $G^\infty(U)$, and define a probability distribution λ on $\mathbb{R}$ by

$$\lambda(A) = \frac{1-\gamma}{2\beta}\, \mathbb{E}\left[\sum_{((0,U),(x,s))\in E(G^\infty(U))} \delta_{\mathtt{d}(0,x)}(A) \right].$$

Another application of Theorem 1 enables us to prove convergence of λ_t to λ, in probability. It is of course not possible to have an explicit expression for λ. However, we can estimate its tail behaviour in the case of a polynomially decaying profile function.

Theorem 5. *Suppose that there exists $\delta > d$ such that the profile function satisfies $\varphi(x) \asymp (1+x)^{-\delta}$. Then*

$$\lambda([K, +\infty)) \asymp (1+K)^{-\eta},$$

where $\eta \in (0, d]$ is the smallest of the three constants d, $d(\frac{1}{\gamma} - 1)$ and $\delta - d$.

The proof is the most technical of our work and is omitted here. Note that if $d = 1$ or if $\gamma \geq \frac{d}{1+d}$ or if $\delta \leq d + 1$, then λ does not have a first moment, and the *mean* edge length is not of order $t^{-1/d}$. The heavy tails of the empirical edge length distribution highlight the nature of our networks as *small worlds*.

The empirical edge length distribution for the original SPA model, corresponding roughly to the case $\delta = \infty$, is also studied in Janssen et al. [15]. They show that if $\gamma > \frac{1}{2}$ and $\frac{3\gamma+2}{4\gamma+2} < \alpha < 1$, then

$$\left|\left\{\text{edges of length longer than } t^{-\alpha/d}\right\}\right| \sim C\, t^{(2-\alpha)+\frac{1}{\gamma}(\alpha-1)}$$

for an explicit constant $C > 0$. Our result uses a different order of limits, but leads to the same order of growth for the comparable quantity $t\lambda[t^{(1-\alpha)/d}, \infty)$. Note that the general form of the profile functions allows for a genuinely richer phenomenology in our case.

5 Further work

In [8], the authors find small separators for the SPA model. They deduce that the spectral gap of the normalised Laplacian of the graph G_t converges to 1, yielding bad expansion properties for G_t. The separators they found are simply obtained by cutting the hypercube in half. We expect that the same strategy would yield similar results for our generalised model, with the slight difference that the separators will not be as small, depending on the tail of the profile function φ.

Existence of a giant component. Let us forget about the orientation of the edges of G_t, and simply consider it as an unoriented graph. Note that, as $\mu * \nu(0) > 0$, the graph has a number of isolated vertices growing linearly in time, and is therefore not connected. Before using G_t as a model for the world wide web, we would like to ensure the existence of a giant component of G_t, i.e. a connected component of linear size. Moreover, we would be interested in the robustness of a giant component under random attack. Robustness is defined by the existence, for every positive value of p, of a giant component in the graph obtained by removing every edge independently with probability $1 - p$.

Proposition 1 suggests that the existence of a giant component for G_t should be related to percolation in G^∞, that is, the existence of an *infinite* connected component in G^∞. As the construction of the graph G^∞ is based (at least in the case of an indicator profile functions) on balls with random positions and random sizes, it resembles the construction of random geometric graphs, and so it seems plausible to use methods from this field as surveyed, for example, in Meester and Roy [17]. Just like in continuum percolation, we cannot really hope for a precise criterion deciding whether there is or is no percolation in G^∞ for any attachment rule f and any profile function φ. However, we hope to identify the domain of existence of a robust giant component. At this point we conjecture the following results, based on preliminary calculations, which highlight interesting phase transitions not occurring for non-spatial models, and show the crucial role of the tail of the profile function (at least in dimension 1). The phase corresponding to our first conjecture seems the best candidate for a model of the world wide web. In (1) and (3) we assume the profile function satisfies

$$\varphi(x) \asymp (1+x)^{-\delta}.$$

(1) *There is always a robust giant component if* $\gamma > 1 - \frac{1}{1+\delta/d}$.

In this case, the shortest paths between two typical vertices in the giant component is doubly logarithmic in the number of nodes. Similarly to the non-spatial models there is a 'core' of old vertices connected to each other in no more than a finite number of steps. Short paths between remote vertices typically connect via this core. The condition on γ, and in particular a finite value of δ, is necessary for the formation of a core.

(2) *There is never a robust giant component if* $\gamma < 1/2$.

This conjecture is based on the corresponding result for non-spatial preferential attachment networks, and the idea that the spatial correlations of the model cannot help the construction of a giant component.

(3) *There is never a robust giant component if* $d = 1$ *and* $\gamma < 1 - 1/\delta$.

Here we have a strong concentration of the length of vertices around the typical value, which give the network some characteristics of geometric random graphs.

All the other cases seem to be even tougher open questions. It would be interesting if we could get a robust giant component in a case not covered by our first conjecture, as this giant component would then have a very different topology from the one in the non-spatial models.

Acknowledgments. We gratefully acknowledge support of this project by the *European Science Foundation* through the research network *Random Geometry of Large Interacting Systems and Statistical Physics (RGLIS).*

References

1. W. Aiello, A. Bonato, C. Cooper, J. Janssen, and P. Prałat. A spatial web graph model with local influence regions. *Internet Mathematics*, 5:175–196, 2009.
2. R. Albert and A.-L. Barabási. Statistical mechanics of complex networks. *Rev. Modern Phys.*, 74:47–97, 2002.
3. A.-L. Barabási and R. Albert. Emergence of scaling in random networks. *Science*, 286:509–512, 1999.
4. N. Berger, C. Borgs, J. Chayes, and A. Saberi. Asymptotic behavior and distributional limits of preferential attachment graphs. *Ann. Prob.*, to appear, 2013.
5. M. Bloznelis. Degree and clustering coefficient in sparse random intersection graphs. *Ann. Appl.Prob.*, 23:1254–1289, 2013.
6. B. Bollobás and O. M. Riordan. Mathematical results on scale-free random graphs. In *Handbook of graphs and networks*, pages 1–34. Wiley-VCH, Weinheim, 2003.
7. A. Bonato, J. Janssen, and P. Pralat. Geometric protean graphs. *Internet Mathematics*, 8:2–28, 2012.
8. C. Cooper, A. Frieze, and P. Pralat. Some typical properties of the spatial preferred attachment model. In *Proceedings of the 9th Workshop on Algorithms and Models for the Web Graph (WAW 2012)*, volume 7323 of *Lecture Notes in Comput. Sci.*, pages 29–40. Springer, Berlin, 2012.
9. S. Dereich and P. Mörters. Random networks with sublinear preferential attachment: degree evolutions. *Electron. J. Probab.*, 14:no. 43, 1222–1267, 2009.
10. S. Dereich and P. Mörters. Random networks with concave preferential attachment rule. *Jahresber. Dtsch. Math.-Ver.*, 113:21–40, 2011.
11. S. Dereich and P. Mörters. Random networks with sublinear preferential attachment: the giant component. *Ann. Prob.*, 41:329–384, 2013.
12. A. D. Flaxman, A. M. Frieze, and J. Vera. A geometric preferential attachment model of networks. *Internet Math.*, 3:187–205, 2006.
13. A. D. Flaxman, A. M. Frieze, and J. Vera. A geometric preferential attachment model of networks. II. In *Algorithms and models for the web-graph*, volume 4863 of *Lecture Notes in Comput. Sci.*, pages 41–55. Springer, Berlin, 2007.
14. E. Jacob and P. Mörters. Spatial preferential attachment networks: Power laws and clustering coefficients. *arXiv:1210.3830*, 2012.
15. J. Janssen, P. Pralat, and R. Wilson. Geometric graph properties of the spatial preferred attachment model. *Adv. Appl. Math.*, 50:243–267, 2013.
16. J. Jordan. Degree sequences of geometric preferential attachment graphs. *Adv. in Appl. Probab.*, 42:319–330, 2010.
17. R. Meester and R. Roy. *Continuum percolation*, volume 119 of *Cambridge Tracts in Mathematics*. Cambridge University Press, Cambridge, 1996.
18. M. Penrose. *Random geometric graphs*, volume 5 of *Oxford Studies in Probability*. Oxford University Press, Oxford, 2003.
19. M. D. Penrose and J. E. Yukich. Weak laws of large numbers in geometric probability. *Ann. Appl. Probab.*, 13:277–303, 2003.
20. A. Rudas, B. Tóth, and B. Valkó. Random trees and general branching processes. *Random Structures Algorithms*, 31:186–202, 2007.